Case mix models using prior dose-response variable

MMA §623e: ESRD Bundled Payment Demonstration

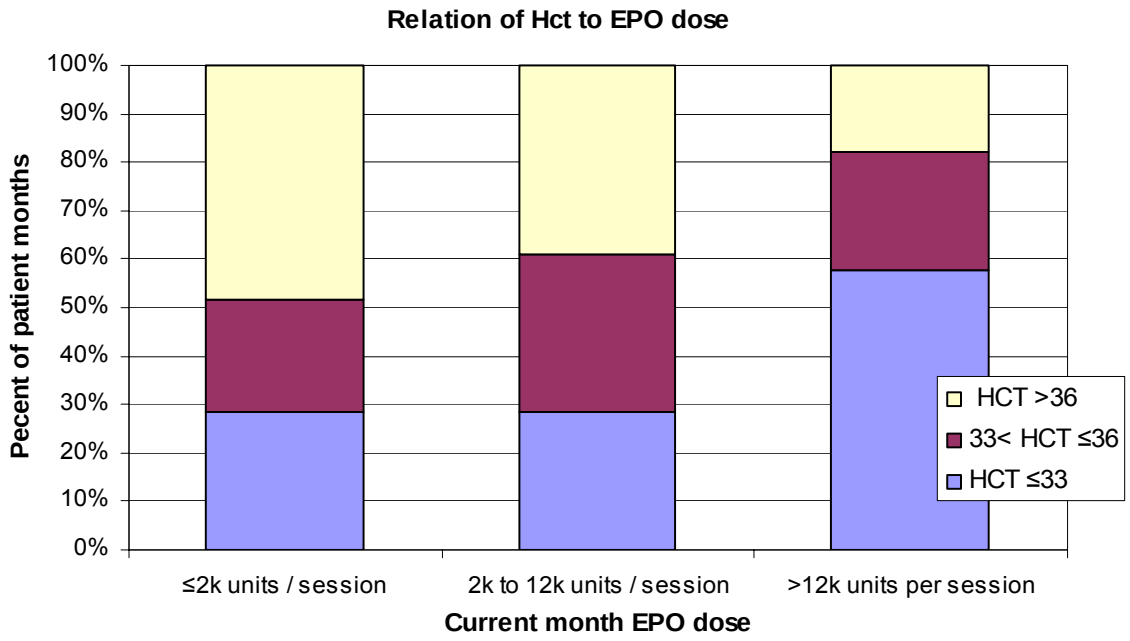
The initial case mix modeling effort demonstrated the limited ability of patient characteristics such as diagnosis, age, socio-economic and demographic characteristics to strongly predict resource use. Because EPO dominates all bundles of separately billed items and services, the problem of predicting resource use will depend substantially on the ability of a model to predict use of EPO. These data were consistent with information from clinical trials and evidence-based practice guidelines that document substantial patient-to-patient variation response to EPO.

Available evidence suggests that the best predictor of both EPO dose and EPO response for the current month is prior EPO dose and prior EPO response. This should not be surprising. The consensus of professional opinion (e.g., from K/DOQI guidelines and the USRDS 2004 annual data report) appears to be that EPO use is generally consistent with dosing guidelines and has been responsible for a significant improvement in anemia outcomes. The evidence from clinical trials related to EPO also supports the conclusion that intrinsic variation in biological/physiological response to EPO produces wide variation in the 'effective' dose required to maintain hemoglobin levels in the target range.

USE OF EPO IN RELATION TO HEMATOCRIT

Figure 1 presents the results of a simple cross-tabulation of prior month's hematocrit against current month's EPO dose (measured in units per session).

Figure 1



Note: Each bar displays the distribution of patients receiving a dose of EPO in the specified range across the specified range of hematocrit values.

Patients receiving a high dose of EPO are more likely to have a low hematocrit than patient receiving a lower dose: 58 percent of patients receiving more than 12,000 units of EPO per session (more than 36,000 units per week) have an hematocrit less than or equal to 33. In contrast, 48 percent of patients receiving up to 2,000 units of EPO per session (6,000 units per week) have an hematocrit greater than 36.

These data suggest that use of EPO is generally consistent with dosing guidelines or, at least, that hematocrit is a determinant of EPO dose. A further implication is that the best predictor of current use will be some combination of prior use and an indicator of response. The simplest 'prior use' model would simply express the current month's dose as a simple function of last month's dose. This model would create incentives that are similar to those created by FFS payment. It makes no reference to hemoglobin levels and gives 'full credit' for EPO use even if a patient has a high hemoglobin level. The EPO dosing guidelines (and K/DOQI practice guidelines) essentially call for EPO dose to be reduced when the patient's hemoglobin is above the upper limit of the target range.

PROPOSED PRIOR DOSE-RESPONSE VARIABLE

If a principal determinant of EPO use is prior response to EPO therapy, then it is possible that response to EPO at an earlier point in time may be a powerful predictor of current (or future) use of EPO and, hence, of the total use of separately billed services by dialysis patients. To evaluate this possibility, a dose-reponse variable was constructed. Several ways of constructing such a variable can be identified, but the simplest simply divides a patient's prior period's EPO dose (or other measure of usage) by the patient's hematocrit. This variable is consistent, in some ways, with the dosing guidelines for EPO, which call for reducing dose as hematocrit or hemoglobin rises.

The dose response variable can be challenging to interpret. By translating hematocrit values into hemoglobin values the variable can be more readily interpreted as the number of units of EPO (per session) administered per g/dL of hemoglobin. Table 1 provides several illustrative examples of the proposed EPO dose-response variable over a range of combinations of dose and hematocrit.

Table 1: Illustrations of prior EPO dose-response variable

		EPO Dose (Units)			% difference from E	
Patient	Hct	per session	per week	Dose-response	Dose	Dose-Response
A	31	9,000	27,000	290	80%	103%
B	32	8,000	24,000	250	60%	75%
C	33	7,000	21,000	212	40%	48%
D	34	6,000	18,000	176	20%	24%
E	**35**	**5,000**	**15,000**	**143**	**0%**	**0%**
F	36	4,000	12,000	111	-20%	-22%
G	37	3,000	9,000	81	-40%	-43%
H	38	2,000	6,000	53	-60%	-63%
I	39	1,000	3,000	26	-80%	-82%

At a given dose of EPO, the amount of credit that the model will give in predicting future EPO use declines as the patient's hematocrit increases and increases as the patient's hematocrit falls. A patient with a high hematocrit receives less 'credit' for prior EPO use than a patient with a low hematocrit.

In Table 1, these relationships are illustrated by calculating changes in the amount of 'credit' that the model gives for prior use relative to Patient E. Patient A can be used to illustrate how the dose response variable will change in response to changes in hematocrit.

- Patient A has an hematocrit that is 11 percent lower than patient E's and receives a greater dose of EPO that is 80 percent greater. The value of the dose-response variable for Patient E is 103 percent higher than that of patient E.
- If Patient A's hematocrit rises to 33, and the dose continues at 9,000 units, the value of the dose-response variable will fall but will still be 91 percent higher than that of Patient E even though the dose of EPO is only 80 percent higher.
- If Patient A's hemtocrit continues to rise with no reduction in dose, the value of the dose-response variable will continue to fall. When Patient A's hematocrit reaches 36 the value of the dose-response variable will be 75 percent higher than that of Patient E.

The precise implications of this variable for the incentives that are created undoubtedly merit and will receive extensive discussion.

INITIAL MODELS USING PRIOR DOSE-RESPONSE

To evaluate the potential ability of the prior dose-response measure to predict current resource use, five models were estimated. All models predicted the use of services included in bundle 1C. (That is, the models did not predict EPO use alone. Instead, EPO use was used to predict total resource use. However, EPO by itself represents 60 percent of the total Medicare allowable charges included in bundle 1C; EPO and related drugs used in the management of anemia account for nearly 90 percent of bundle 1C. As a result, the problem of predicting resource use is substantially a problem of predicting EPO use.)

These models used either of two distinct dose-response variables. One variable measured dose relative to hematocrit over the three months prior to the 'current' month (i.e., the month whose resource use was being predicted). The other variable measured dose relative to hematocrit over the sixth, seventh and eighth months prior to the 'current' month.

The additional variables used in some models included:

- Hematocrit in the prior month.

- Hematocrit at the start of dialysis.
- Patient age, sex, and socio-demographic characteristics (race and ethnicity).
- Duration of renal replacement therapy.
- Measures of body size (body surface area and low body mass).
- Selected behavioral health measures and measures of functional status.
- The presence or absence of diagnoses for 36 categories of clinical conditions including cardio-vascular disease, infectious disease, cancer, anemia and conditions related to anemia, diabetes, and selected other conditions.
- Recent admission for inpatient hospital care.

Some models also included a series of variables to control or test for 'fixed' facility effects.

Table 2 presents summary statistics on the performance of these models. In the table the "base case mix (CM) variables" refers to all variables *except* prior dose-response, previous month's hematocrit, and the 'fixed' facility effects variables.

Effect on predictive power

The 'base' model (Model 1) included only the case mix variables that were used in the models presented to the Advisory Board in May. This model had an R^2 of 0.074. That is, it 'predicted' or accounted for only 7.4% of the month-to-month variation in total resource use among patients.

Model 2 uses the prior dose-response for the 'recent' period and is much more able to predict current resource use (R^2 of 0.759). That is, it accounts for more than three quarters of the patient-to-patient variation in resource use per session.

Model 4 uses the dose-response in the more 'recent' period plus the prior month's hematocrit and achieves an R^2 of 0.887. This model accounts for nearly 90 percent of patient-to-patient variation in resource use per session. In other words, it comes close to reflecting actual resource use.

Model 3 uses the dose-response in a more 'distant' period and has an R^2 of 0.279. It accounts for more than a quarter of patient-to-patient variation in resource use, which is nearly four times as much of the variation that the basic case mix variables alone were able to predict. The dose-response variable for the more distant period accounts for less than half as much of the patient-to-patient variation as the dose-response variable for the more recent three month period.

(Model 5, which uses the more 'recent' dose-response measure and accounts for differences in resource use between facilities, accounts for slightly less patient-to-patient variation than the model using the prior month's hematocrit.)

Consistent with the higher R^2 values, the models that include prior dose-response predict greater variation in resource use and leave less variation unexplained.

- Model 1 uses patient characteristics only but not the dose-response variables, and predicts variation in resource use that is about one quarter as large as actual variation. The standard deviation of predicted resource use is nearly $60 less than the standard deviation of actual resource use.
- The models using the 'recent' dose-response variable, predict variation that is nearly as large as actual variation. The standard deviation of predicted resource use is within $5 to $10 of the standard deviation of actual resource use.
- Model 3, which uses the dose-response variable for the more distant period, predicts variation that is about half that of actual variation. The standard deviation of predicted resource use is nearly $40 less than the standard deviation of actual resource use.

Table 2: Performance statistics for initial prior use models

	Model 1	Model 2	Model 3	Model 4	Model 5
Model form	Log	Log	Log	Log	Log
Base CM model	X	X	X	X	X
Prior EPO Response (1 to 3 months)**		X		X	X
Prior EPO Response (6 to 8 months)**			X		
Hct (prior month)				X	
Facility					X
R-squared	0.074	0.759	0.279	0.887	0.818
Mean (Actual) MAC	$101.37	$101.37	$101.37	$101.37	$101.37
Std. Dev. Predicted	$22.24	$71.08	$43.14	$76.87	$73.78
Std. Dev. Actual	$81.60	$81.60	$81.60	$81.60	$81.60
Correlation with prior month prediction error					
t-1	0.73	0.49	0.66	0.51	0.47
t-2	0.62	0.29	0.49	0.37	0.25
t-3	0.57	0.20	0.38	0.29	0.16
t-4	0.53	0.22	0.29	0.28	0.18
t-5	0.50	0.23	0.21	0.27	0.18
t-6	0.48	0.21	0.12	0.26	0.17
t-7	0.45	0.19	0.08	0.23	0.15
t-8	0.43	0.18	0.08	0.22	0.14
t-9	0.42	0.18	0.13	0.22	0.14
t-10	0.40	0.16	0.13	0.20	0.13
t-11	0.38	0.15	0.12	0.18	0.11

Effect on correlation of prediction errors over time

The base model showed a high correlation in prediction errors across months for the same patient. This correlation suggests that patients using more resources than predicted in the current month will probably continue to use more resources than predicted in future months. Conversely, patients using fewer resources than predicted in the current month are likely to use fewer resources than predicted in future months as well. Gains or losses are likely to persist, probably because characteristics responsible for use of resources are not included in the model. This situation is consistent with the clinical guidelines on EPO use and evidence from clinical trials, which as noted earlier document large variation in patients' response to EPO.

The correlation of prediction errors over time is affected, as would be expected, by the inclusion of the dose-response variable in the model. Models 2 and 4 using the 'recent' dose-response variable sharply reduced the correlation of prediction errors in months close to the 'current' months.

Model 4, using the more 'distant' dose-response variable, did not reduce the correlation of prediction errors in the months close to the 'current' month by as much as the model using the more 'recent' dose-response variable, but resulted in a larger reduction in the correlation of prediction errors in later months.

It should be noted that the correlation of prior months prediction errors is not likely to be zero. The resource use by patients of a 'high cost' facility will (even with 'perfect' case mix adjustment) be greater than predicted, and that difference will persist over time. As a result, some correlation of patients' prediction errors over time is to be expected. However, the effect of differences in resource use among facilities on correlation of prediction errors over time is not easily distinguished from the effect of leaving out of the model patient characteristics that are predictive of resource use.

Effect on case mix variables

The inclusion of prior dose-response has a substantial effect on the estimated relationship between other case mix variables (i.e., patient characteristics other than use of EPO and hematocrit) and use of resources. When prior dose-response is included, a much smaller number of characteristics are shown to have be associated with differences in resource use that are larger than ±5 percent. However, some characteristics show a larger relationship. For example, African American's had 6 percent greater resource use in the model that does not include prior dose-response, but 13 percent greater resource use in the models that include the prior dose-response.

The more 'recent' dose-response measure had a stronger effect on the other case mix variables than the more 'distant' measure. In other words, when the use of resources in the recent past is used to predict resource use in the current month, a larger number of the remaining case mix variables become relatively unimportant. These variables continued to be more important in accounting for variation in

resource use when the more 'distant' dose response measure was used. In both cases, the large number of observations used in these analysis means that most of the case mix variables continue to have a statistically significant relationship to resource use.

REFINED MODELS USING PRIOR DOSE-RESPONSE

Based on these initial results, a further set of models were estimated. These models all use the dose-response measure for the more 'distant' period. They also use a smaller number of basic case mix variables than models 1 through 5 that were selected because they showed an impact on resource use greater than ±5 percent in the initial models. This 'reduced' case mix model includes:

- A single variable for patient age (whether the patient is <18 years of age) and a variable for female
- Two variables for racial groups (African American and Pacific Islander)
- A total of 18 diagnostic categories
- Hematocrit at start of dialysis
- Two body size measures (body surface area and low body mass index)
- Hospitalization in the prior month

The duration of renal replacement therapy (vintage) was treated as a separate category of variable in one of these models.

Table 3 summarizes the ability of these models to predict actual resource use. The reduced set of case mix variables does not appear to reduce the ability of the model to predict resource use, particular when the prior dose-response variable is included.

The prior month's hematocrit continues to be a significant predictor of current month resource use. It increases the R^2 by more than a third, from 28 percent (Model 9) to 37 percent (Model 10). As in the earlier models, the prior month's hematocrit is inversely related to the current month's resource use: a higher hematocrit in the prior month is associated with lower resource use in the current month and vice versa.

The correlation of the current month's prediction error with prediction errors for prior months is broadly similar for all models.

An examination of the 'vintage' variables shows that the initial months of dialysis are associated with substantially higher resource use. The initial month of dialysis appears to have resource use that is nearly 60 percent higher than experienced by patients who have received dialysis for more than a year. The first three full months of dialysis (after the initial month) appear to have resource use than is between 25 percent and 35 percent higher than patients who have been receiving dialysis for

more than a year. (This information can be found in the detailed results for the models contained in tab 7 beginning on page 12.)

Table 3:

	Model 6	Model 7	Model 8	Model 9	Model 10
Reduced base CM model	X			X	X
Vintage	X		X	X	X
Prior EPO Response (1 to 3 months)**					
Prior EPO Response (6 to 8 months)**		X	X	X	X
Hct (prior month)					X
Facility					
Patient months	1,944,012	1,944,012	1,944,012	1,944,012	1,944,012
Percent of total used in analysis	79%	79%	79%	79%	79%
R-squared	0.069	0.242	0.242	0.279	0.370
Mean (Actual)	$101.37	$101.37	$101.37	$101.37	$101.37
Std. Dev. Predicted	$21.43	$40.14	$40.10	$43.10	$49.65
Std. Dev. Actual	$81.59	$81.59	$81.59	$81.59	$81.59
Correlation with prior month prediction error					
t-1	0.73	0.68	0.68	0.66	0.64
t-2	0.63	0.51	0.51	0.49	0.50
t-3	0.57	0.39	0.39	0.38	0.39
t-4	0.53	0.29	0.30	0.29	0.29
t-5	0.51	0.20	0.21	0.21	0.22
t-6	0.49	0.10	0.11	0.12	0.14
t-7	0.46	0.06	0.07	0.08	0.11
t-8	0.44	0.06	0.08	0.08	0.12
t-9	0.43	0.11	0.12	0.13	0.15
t-10	0.41	0.12	0.13	0.13	0.15
t-11	0.39	0.12	0.13	0.13	0.14

An examination of the 'diagnosis' variables included in the model has a number of implications for a final case mix model. First, with the exception of septicemia and shock, these categories have low prevalence rates. These conditions appear to be independent of one another, although that may reflect limitations of the available data. The underreporting of clinical data has significant implications for the inclusion of adjustments for dialysis in the payment model.

The magnitude of the relationship between these conditions and resource use is variable, but is generally in the range of 5 percent to 10 percent. When prior month's hematocrit is included in the model only three conditions are estimated as having an effect on resource use that exceeds ±10 percent: GI tract bleeding, myelodysplastic syndrome, and myelofibrosis.

It is possible that a model that merely distinguishes between uncomplicated patients, patients with moderate complications or co-morbid conditions, and patients with major complications or co-morbidity will perform as well as the model that includes separate categories for discrete groups of related diagnoses. From a statistical perspective it is unlikely that such a further simplification of the case mix model will materially erode the ability of the model to predict resource use. From an administrative perspective, it is likely that such the simpler model will reduce the need for adjustments to 'correct' for improvements or changes in the coding of diagnosis on dialysis facility claims. Additional models are being estimated to evaluate the first of these hypotheses.

Effects on facilities

Up to this point, the discussion has focused on the ability of the case mix models to account for variation in resource use among patients. Information on the ability of a model to predict variation in individual patients' use of resources is useful, but it does not directly address the question of the impact that a case mix adjusted payment system will have on facilities. Do any of the statistical models do a "good enough" job of predicting or accounting for variation in resource use at the level of the facility?

In the context of the bundled payment demonstration "good enough" might mean several things. First, it might mean that differences between actual and predicted payment are small enough, infrequent enough, or random enough to suggest that a bundled payment will not have an adverse effect on quality of care, will not expose facilities to an unacceptable risk on financial losses, and will not create risks associated with patient selection.

Second, 'good enough' means that the difference between actual and predicted payment does not vary systematically across types of facilities: large vs. small, urban vs. rural, or by organizational structure.

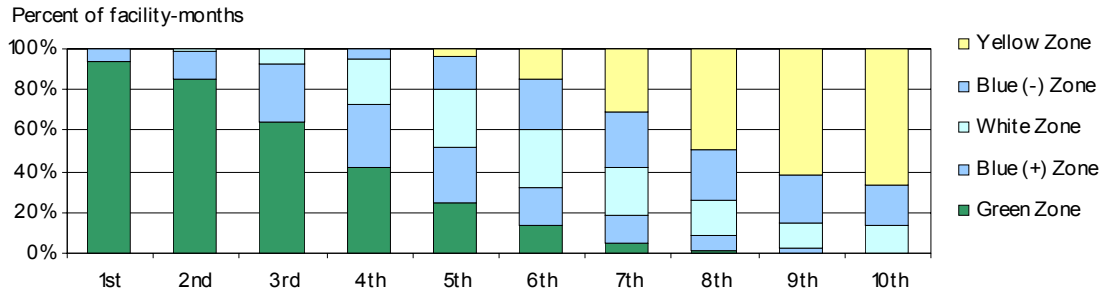
To evaluate whether a case mix model over- or under-predicts payment at the level of the facility, an impact table was constructed comparing actual to predicted resource use. To construct this table, the average Medicare allowable charge (MAC) per session was calculated for each facility month by grouping patient months based on the facility at which the patient was treated. (That is, for each month the average per session MAC for bundle 1C was calculated for each facility's patients.) Facility-months were then divided into deciles; that is, each facility-month was assigned to one of ten 'resource use' categories, ranked from low to high, based on its average MAC per session. The average predicted MAC was then calculated for each facility-month, and the month was assigned to one of the ten 'resource use' categories based on its average predicted MAC per session. The data were then tabulated to identify whether the predicted payment for a facility-month was more-or-less the same as the actual payment, was somewhat higher or lower than actual payment, or was substantially higher or lower than actual payment.

The same analysis was performed for facility-years (i.e., for facilities) with generally similar results. In other words, the impact for a given facility in a single month tends to predict what the impact on the facility will be for the entire year.

Figure 2 summarizes this facility-month analysis and depicts the ability of case mix model 9 to predict resource use at the level of the facility. Model 9 was described in Table 3. This figure describes only the use of separately billed items and services included in bundle 1C. Each column in the figure represents 10 percent of facility-months. The first column represents the months with the lowest resource use (up to $52 per session). The tenth column represents the months with the highest resource use ($160 per session and above). Information on the specific ranges covered by the ten 'resource use' categories can be found on page 5 of the material included in Tab 8.

The portion of each bar in Figure 2 designated as representing the 'green zone' represents those months in which the predicted resource use was two deciles or more *higher* than actual resource use. These are months during which the facility would experience a surplus. The portion of each bar designated as representing the 'yellow zone' represents those months in which the predicted resource use was two deciles or more *lower* than the actual resource use. These are months during which the facility would experience a loss. The 'blue zone' designates months in which average predicted resource use places the facility one decile higher (blue +) or below (blue -) actual resource use. The 'white zone' identifies months in which the predicted and actual resource used placed the facility in the same decile. The more 'accurately' a model predicts resource use, the higher the percentage of months that will fall into the blue or white zones.

Figure 2: Facility Impact Summary for Model 9



Summary	Months	Percent	
Same decile	7,401	15%	White zone
1 decile higher or lower	14,226	29%	Blue zone
2 or more deciles higher	16,104	33%	Green zone
2 or more deciles lower	11,105	23%	Yellow zone
Total	48,836	100%	

Predicted resource use tends to be higher than actual resource use in facility months during which resource use is low or below average. At the facility level, this means

that predicted resource use will tend to be higher than actual resource use in facilities with low or below average resource use. In other words, these facilities are likely to experience a net gain from implementation of a case mix adjusted payment system.

The opposite is true of facility months in which resource use is high or above average. These facilities are likely to experience losses.

In the middle of the range, from the 4th to the 7th deciles, between 58 percent and 72 percent of months had predicted resource use that was close to actual resource use, i.e., in the 'blue' or 'white' zones.

Model 9, which explained 'only' 28 percent of patient-to patient variation in resource use, does generate average predicted values for facilities that cover a substantial portion of the range of actual values. It appears to do a better job of predicting high resource use than low resource use. A third of all facility months have average predicted values that are substantially higher than the average actual values. Almost one quarter of all facility months have predicted values that are substantially below the average actual values.

PRELIMINARY CONCLUSIONS AND QUESTIONS FOR DISUCUSSION

1. Prior dose-response is likely to be needed in any case mix model. The selection of a period to be covered by the dose-response variable is not obvious, however. Advice is needed on the choice of an appropriate period. Several alternatives can, obviously, be considered: (1) the three months prior to the current month; (2) a more 'distant' period (e.g., the 6th, 7th, and 8th months prior to the current month); and (3) a somewhat longer period immediately adjacent to the prior months (e.g., the 4 or 5 months prior to the current month).

2. Using a dose-response variable has administrative implications for the collection of information on claims forms and for the availability and reporting of information for patients who change facilities. It also has implications for the nature of the incentives that are created. For example, as structured in these models, an increase in a patient's hematocrit will in and of itself cause a reduction in future payment. However, if dose is reduced in response to the higher hematocrit future payment will be further reduced. This simplistic analysis, however, considers only the effect of changes in dose on payment. It does not consider the effect of changes in dose on costs or net income.

3. The prior month's hematocrit appears to be an important predictor of resource use. It also may be important clinically because it is the hematocrit in the current month that affects dose adjustments. However, including it in a case mix adjustment could create incentives to reduce EPO dosage because a lower hematocrit results in higher payment. Additional discussion of this issue and a recommendation is needed.

4. An adjustment is likely to be needed for the initial months of dialysis. However, the period of time that should be covered by this adjustment is unclear. It is

clearly longer than one month, and probably shorter than 12. Input on the structure of such and adjustment would be useful. For example, would separate adjustments for the initial month of dialysis and for the first three full months of dialysis be seen as administratively viable? The length of this period may depend on the length of the period over which dose-response is measured.

5. The way in which the diagnosis variables should be included in the model is unclear. In the near term, the approach that makes the most sense on technical and statistical grounds is to distinguish between complicated and non-complicated patients or, possibly, between patients with no complications, moderate complications, and severe complications. Over the longer run, it may be appropriate to develop a more refined set of diagnostic categories. Would such an approach be acceptable to clinicians and to facilities?